



# СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ



## «АНАЛІЗ ДАНИХ»

Рівень освіти	Бакалавр	Тип дисципліни	Вибіркова. Професійна
Шифр та назва спеціальності	122 – Комп'ютерні науки	Інститут	ННІ КНІТ Навчально науковий інститут комп'ютерних наук та інформаційних технологій
Назва освітньо-наукової програми	Комп'ютерні науки	Кафедра	Системного аналізу та інформаційно-аналітичних технологій

### ВИКЛАДАЧ

**Безменов Микола Іванович, [mykola.bezmenov@kpi.edu.ua](mailto:mykola.bezmenov@kpi.edu.ua)**

Кандидат технічних наук, доцент, професор НТУ «ХПІ», професор кафедри системного аналізу та інформаційно-аналітичних технологій НТУ «ХПІ». Досвід роботи – більш 40 років. Автор понад 100 наукових та навчально-методичних праць. Провідний лектор з дисциплін: «Алгоритмізація та програмування», «Інформатика і програмування», «Основи візуального програмування», «Аналіз даних», «Статистичний аналіз даних», «Статистична обробка соціально-економічної інформації»

Персональна сторінка - <https://web.kpi.kharkov.ua/say/uk/uaabout/uaprofs/bezmenovmi/>

### ЗАГАЛЬНА ІНФОРМАЦІЯ ПРО ДИСЦИПЛІНУ

Анотація	Дисципліна спрямована на оволодіння теоретичних основ статистичного аналізу даних. Розглядаються питання аналізу зв'язків між параметрами, що описують складні системи, методи відсіювання помилкових даних, методи кластерного і факторного аналізу. У рамках курсу розглядаються питання використання апарату теорії ймовірностей та математичної статистики для статистичного аналізу експериментальних даних для подальшої побудови математичних моделей складних систем.
Мета та цілі	Мета вивчення дисципліни: освоєння підходів та засобів аналізу емпіричних даних за допомогою теоретичних та евристичних методів для подальшого використання результатів аналізу при побудові математичних моделей складних технічних і соціально економічних систем.
Формат	Лекції, лабораторні роботи, консультації. Підсумковий контроль - іспит
Результати навчання	Володіти навичками розробки програмного забезпечення для вирішення задач аналізу даних, методами та засобами оцінювання якості та тестування програмного забезпечення.
Обсяг	Загальний обсяг дисципліни 150 год.: лекції – 32 год., лабораторні роботи – 32 год., самостійна робота – 86 год.
Пререквізити	«Алгебра та геометрія», «Фізика», «Теорія ймовірностей та математична статистика»

**Вимоги викладача**

Студент зобов'язаний відвідувати всі заняття згідно розкладу, не спізнюватися. Дотримуватися етики поведінки. Працювати з навчальної та додатковою літературою, з літературою на електронних носіях і в Інтернеті. При пропуску лекційних занять проводиться усна співбесіда за темою. Відпрацьовувати лабораторні заняття при наявності допуску викладача. З метою оволодіння необхідною якістю освіти з дисципліни потрібна відвідуваність і регулярна підготовленість до занять. Без особистої присутності студента підсумковий контроль не проводиться.

**СТРУКТУРА ДИСЦИПЛІНИ**

<b>Лекція 1</b>	Предмет і задачі дисципліни її місце у навчальній програмі. Дані й виміри. Типи даних і шкал	<b>Лабораторна робота 1</b>	Перевірка гіпотез про ймовірність для дискретних параметрів	<b>Самостійна робота</b>	Освоєння середовища MathCAD
<b>Лекція 2</b>	Аналіз дихотомічних спостережень з використанням біноміального розподілу. Аналіз дихотомічних спостережень з використанням z-статистики. Перевірка гіпотез про приналежність одному з декількох класів. Використання статистики $\chi^2$ у випадку двох класів	<b>Лабораторна робота 2</b>	Перевірка гіпотез про ймовірність для дискретних параметрів		Використання статистичних функцій у MathCAD
<b>Лекція 3</b>	Особливості перевірки гіпотез у випадку неперервних змінних. Критерій узгодження $\chi^2$ . Критерій узгодження Колмогорова – Смірнова. Перевірка гіпотез і довірчій інтервал для $\mu$ . Перевірка гіпотез і довірчій інтервал для $\sigma^2$	<b>Лабораторна робота 3</b>	Перевірка гіпотези про відповідність розподілу неперервного параметра деякому закону		Програмування циклічних процесів у MathCAD
		<b>Лабораторна робота 4</b>	Перевірка гіпотез про математичне очікування і дисперсії		
<b>Лекція 4</b>	Аналіз двох випадкових величин. Дві випадкові величини з однієї сукупності. Аналіз однієї випадкової величини на двох сукупностях. Аналіз K випадкових величин на одній сукупності. Аналіз однієї випадкової величини на K сукупностях	<b>Лабораторна робота 5</b>	Аналіз двох випадкових величин із однієї сукупності		Обчислення асимптотичних стандартних помилок для мір зв'язку
		<b>Лабораторна робота 6</b>	Аналіз випадкової величини на двох сукупностях		
<b>Лекція 5</b>	Аналіз таблиць спряженості ознак	<b>Лабораторна робота 7</b>	Перевірка гіпотези про незалежність двох дискретних параметрів		Методи нормування змінних
<b>Лекція 6</b>	Міри зв'язності Гудмена – Краскала	<b>Лабораторна робота 8</b>	Міри зв'язності Гудмена – Краскала		
<b>Лекція 7</b>	Обробка аномальних спостережень: Вінзоризовані та усічені оцінки. Кусочно-лінійна M-оцінка Гампеля.	<b>Лабораторна робота 9</b>	Оцінювання математичного очікування при наявності викидів у вимірах		
<b>Лекція 8</b>	Аналіз викидів у багатовимірному випадку. Алгоритми обробки даних з наявністю викидів	<b>Лабораторна робота 10</b>	Аналіз викидів у багатовимірному випадку		
<b>Лекція 9</b>	Перевірка гіпотез та класифікація об'єктів засобами багатовимірного аналізу. Баєсівський підхід до класифікації	<b>Лабораторна робота 11</b>	Перевірка гіпотез про вектори середніх у багатовимірному випадку		
<b>Лекція 10</b>	Перевірка гіпотез про вектори середніх при багатовимірному біноміальному і нормальному розподілах	<b>Лабораторна робота 12</b>	Перевірка гіпотез про вектори середніх у багатовимірному випадку		

<b>Лекція 11</b>	Основи дискримінантного аналізу	<b>Лабораторна робота 13</b>	Класифікація багатовимірних об'єктів при наявності інформації про закон розподілу	<b>Самостійна робота</b>
<b>Лекція 12</b>	Покроковий дискримінантний аналіз	<b>Лабораторна робота 14</b>	Класифікація багатовимірних об'єктів при наявності інформації про закон розподілу	
		<b>Лабораторна робота 15</b>	Класифікація багатовимірних об'єктів за допомогою дискримінантних функцій при невідомому розподілі векторів спостережень	
<b>Лекція 13</b>	Поняття ентропії та інформації. Використання інформаційних показників для розв'язання задач класифікації	<b>Лабораторна робота 16</b>	Класифікація багатовимірних об'єктів з використанням інформаційних показників	
<b>Лекція 14</b>	Загальні поняття кластерного аналізу. Гіпотеза компактності. Послідовні та паралельні алгоритми кластерного аналізу			
<b>Лекція 15</b>	Поняття головних компонент. Класифікація за допомогою метода головних компонент.			
<b>Лекція 16</b>	Основні поняття факторного аналізу. Обчислення головних та обертаня факторів.			Методи екстремального групування параметрів Зменшення розмірності факторного простору

## ЛІТЕРАТУРА ТА НАВЧАЛЬНІ МАТЕРІАЛИ

<b>Основна</b>	<ol style="list-style-type: none"> <li>1. Afifi A. A., Azen S. P. Statistical Analysis. Computer Oriented Approach. New; York-San Francisco'; London: Academic Press, 1979. 442 p.</li> <li>2. Карташов М.В. Імовірність, процеси, статистика : посіб. Київ: Видавничо-поліграфічний центр «Київський університет», 2008. 494 с. URL: <a href="https://probability.knu.ua/userfiles/kmv/VPS_Pv.pdf">https://probability.knu.ua/userfiles/kmv/VPS_Pv.pdf</a></li> <li>3. Бахрушин В. Є. Методи аналізу даних : навч. посіб. для студ. Запоріжжя: КПУ, 2011. 268 с.</li> <li>4. Бахрушин В. Є. Аналіз даних : навч. посіб. Запоріжжя: ГУ "ЗІДМУ", 2006. 128 с.</li> <li>5. Руденко В. М. Математична статистика. Київ: Центр учбової літератури, 2012. 304 с.</li> <li>6. Мхітарян В. С. Аналіз даних. URL: <a href="https://stud.com.ua/93298/statistika/analiz_danih">https://stud.com.ua/93298/statistika/analiz_danih</a></li> <li>7. Єлісеєва В. І. Статистика. URL: <a href="https://stud.com.ua/74988/statistika/statistika">https://stud.com.ua/74988/statistika/statistika</a></li> <li>8. Черняк О. І. Інтелектуальний аналіз даних: підруч. Київ: Знання, 2014. 599 с.</li> <li>9. Данильченко О. М., Данильченко А. О. Інтелектуальний аналіз даних: навч. посіб. Житомир: ЖДТУ, 2009. 405 с.</li> <li>10. Ситник В. Ф., Краснюк М. Т. Інтелектуальний аналіз даних (дейтамайнінг): навч. посіб. Київ: КНЕУ, 2007. 376 с.</li> <li>11. Бондаренко Я. С., Кравченко С. В. Посібник до вивчення дисципліни «Статистичний аналіз даних». Дніпро: Ліра, 2018. 40 с. URL: <a href="https://www.researchgate.net/publication/335340316_Posibnik_do_vivcenna_disciplini_Statisticnij_analiz_danih">https://www.researchgate.net/publication/335340316_Posibnik_do_vivcenna_disciplini_Statisticnij_analiz_danih</a></li> </ol>	<b>Додаткова</b>	<ol style="list-style-type: none"> <li>1. Методичні вказівки до лабораторної роботи «Перевірка гіпотез про ймовірність для дискретних параметрів» з курсу «Аналіз даних» для студ. напряму 6.040302 – Інформатика і курсу «Аналіз даних і знань» для студ. напряму 6.040303 – Системний аналіз / Уклад. М. І. Безменов, О. М. Безменова. Харків: НТУ «ХПІ», 2015. 24 с.</li> <li>2. Методичні вказівки до лабораторних робіт за темою «Перевірка гіпотез для неперервних параметрів» з курсу «Аналіз даних» для студ. спеціальностей 122 – Комп'ютерні науки та інформаційні технології, 124 – Системний аналіз, 186 – Видавництво та поліграфія / Уклад. М. І. Безменов, О. М. Безменова, Д. В. Калінін. Харків: НТУ «ХПІ», 2017. 44 с.</li> <li>3. Методичні вказівки до лабораторних робіт за темою «Оцінювання наявності зв'язку між дискретними параметрами» з курсу «Аналіз даних» для студ. спеціальностей 122 – Комп'ютерні науки та інформаційні технології, 124 – Системний аналіз, 186 – Видавництво та поліграфія / Уклад. М. І. Безменов, О. М. Безменова, Д. В. Калінін. Харків: НТУ «ХПІ», 2017. 57 с.</li> <li>4. Методичні вказівки до лабораторних робіт за темою «Обробка аномальних спостережень» з курсу «Аналіз даних» для студ. спеціальностей 122 – Комп'ютерні науки та інформаційні технології, 124 – Системний аналіз, 186 – Видавництво та поліграфія / Уклад. М. І. Безменов, О. М. Безменова, Д. В. Калінін. Харків: НТУ «ХПІ», 2017. 33 с.</li> </ol>
----------------	---	------------------	--

5. Методичні вказівки до лабораторних робіт за темою «Перевірка гіпотез про вектори середніх і класифікація об'єктів для багатовимірного випадку» з курсу «Аналіз даних» для студ. спеціальностей 122 – Комп'ютерні науки та інформаційні технології, 124 – Системний аналіз, 186 – Видавництво та поліграфія / Уклад. М. І. Безменов, О. М. Безменова, Д. В. Калінін. Харків: НТУ «ХПІ», 2017. 64 с.

6. Безменов М. І. Вступ до Mathcad : навч.-метод. посіб. для самот. роб. з курсу «Аналіз даних» для студ. спеціальностей 122 – Комп'ютерні науки, 124 – Системний аналіз, 186 – Видавництво та поліграфія / М. І. Безменов, О. М. Безменова. Харків: НТУ «ХПІ»; ТОВ «Планета-Принт», 2019. 68 с.

## ПЕРЕЛІК ЗАПИТАНЬ ДЛЯ ПІДГОТОВКИ ДО ІСПИТУ

Дані і їхній вимір. Шкали для виміру даних. Перевірка відповідності розподілу деякої теоретичної функції. Перевірка гіпотези про математичне чекання випадкової величини і побудова довірчого інтервалу. Перевірка гіпотези про дисперсії випадкової величини і побудова довірчого інтервалу. Перевірка гіпотези про рівність математичного чекання двох випадкових величин. Перевірка гіпотези про рівність математичних чекань випадкової величини на двох сукупностях. Точний критерій Фішера. Міри зв'язку, засновані на  $\chi^2$ . Критерій  $\chi^2$  для таблиць спряженості і перевірка гіпотез про незалежність. Міри зв'язку для таблиць спряженості довільної розмірності. Міри Гудмена-Краскала. Робастне оцінювання. Перевірка гіпотез про вектори середніх при відомій коваріаційній матриці. Байєсівська процедура класифікації. Емпіричний алгоритм побудови поділяючої функції. Емпіричний алгоритм побудови дискримінантних функцій.

## ПЕРЕЛІК ОБЛАДНАННЯ

Персональні комп'ютери з операційною системою Windows 10, офісними засобами, Visual Studio та Matcad.

## СИСТЕМА ОЦІНЮВАННЯ

Розподіл балів для оцінювання успішності аспіранта	Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	Нарахування балів
	90-100	A	відмінно	
	82-89	B	добре	
	74-81	C		
	64-73	D	задовільно	
	60-63	E		
	35-59	FX		
	0-34	F	незадовільно з обов'язковим повторним вивченням дисципліни	

Бали нараховуються за наступним співвідношенням:

- лабораторні роботи: 20% семестрової оцінки;
- самостійна робота: 20% семестрової оцінки;
- іспит: 60% семестрової оцінки

## НОРМИ АКАДЕМІЧНОЇ ЕТИКИ

Студент повинен дотримуватися «Кодексу етики академічних взаємовідносин та доброчесності НТУ «ХПІ»»: виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність. Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем, а при нерозв'язності конфлікту доводиться до співробітників відділу аспірантури.

Силабус за змістом повністю відповідає робочій програмі навчальної дисципліни

